Linear Algebra and Forbidden Configurations

by

Timothy Scott Pressey

B.A., University of Winnipeg

AN ESSAY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies
(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2017

1 Introduction

We begin with a review of set theory notation. We use the standard notation for the set of natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$ and the set of rational numbers $\mathbb{Q} = \{\frac{a}{b} : a \text{ and } b \text{ have no common factor}\}$. Given a set S, the *cardinality*, or number of elements in a set, is denoted by $|S|$. All sets considered in this essay will be finite. We use the notation $[m] = \{1, 2, \ldots, m\}$. For many other purposes it may be more natural to define $[m]$ to be the set of integers from 0 to $m-1$. In this paper we study matrices, and it is natural to enumerate the rows and columns starting at 1.

The *power set* of a set S, denoted 2^S, is the collection of all subsets of S. This notation is natural as $|2^S| = 2^{|S|}$. In the same vein we will use $\binom{S}{k}$ to represent the family of all k-element subsets of S, and it is pleasing to note that $|\binom{S}{k}| = \binom{|S|}{k}$.

The study of forbidden configurations is a problem in extremal set theory. Classical problems in the field include:

1. What is the largest family of subsets $\mathcal{F} \subseteq 2^{[m]}$ such that no element is contained in another? (Sperner's Theorem for Antichains [**?**])

2. How many k-element pairwise intersecting subsets can we choose from an m-element set? (Erdős-Ko-Rado Theorem [**?**])

The problem of forbidden configurations deals with the largest family of subsets avoiding certain substructures. There are many ways to frame questions of this kind. Early problems of this kind involved finding maximal (or minimal) graphs avoiding (or having) a certain property. As graphs can be thought of as 2-element set systems, this problem generalizes to that of forbidden configurations.

We use the language of $\{0, 1\}$-matrices and configurations in our discussions. A $\{0, 1\}$-*matrix* is a matrix with entries in the set $\{0, 1\}$. Such a matrix is called *simple* if it contains no repeated columns.

Let F be a $\{0, 1\}$-matrix. We will say that a matrix A contains F as a *configuration* if there is a submatrix of A which is a row and column permutation of F and we will use the notation $F \prec A$. There would be alternate definitions including thinking of a configuration as the equivalence class of a matrices which are row and column permutations of F. Since we are manipulating the matrix F repeatedly in this thesis we avoid this equivalence class definition. Another way is to think of A and F as a set systems $\mathcal{A}$, $\mathcal{F}$. In that case $F \prec A$ is the same as saying $\mathcal{A}$ has $\mathcal{F}$ as a *trace*. We will use the notation $||A||$ to denote the number of columns of A, corresponding to $|\mathcal{A}|$, the number of sets in the family $\mathcal{A}$.

Analogous to extremal set problems, we ask for the maximum number of columns in a matrix A on m rows that has no configuration F. To be specific we introduce the following notations

Definition 1.1. *Given a configuration F and an integer m,*

$$Avoid(m, F) = \{A : A \text{ is simple, has } m \text{ rows, and } F \nprec A\}$$

.

Definition 1.2. *Given a configuration F and an integer m,*

$$forb(m, F) = max\{||A|| : A \in Avoid(m, F)\}$$

Hence, our main extremal problem is to compute forb(m, F). Two example calculations are given below. Before we begin, for completeness we define standard terms for manipulating matrices.

A *submatrix* of another matrix is one that can be obtained by removing rows and columns from the original matrix. Note that row and column order is preserved. Given a matrix A and a subset of rows S, we denote $A|_S$ as the submatrix of A obtained after deleting all rows not in S. In most of investigations (not in our discussions of violations in Chapter 4) we are less concerned with row or column order and typically use notations of configurations to describe such situations.

Matrix *concatenation* combines two matrices with the same number of rows. When A and B are two such matrices, we will write $[A|B]$ to be the matrix consisting of the columns of A followed by the columns of B. When t is an integer, the notation $[t \cdot A]$ is used for the concatenation of t copies of A as $[A\,|\,A\,|\,|\cdots|\,A]$.

Example 1. *Calculate forb$(m, [0\ 1])$.*

Solution. Let $A \in$ Avoid$(m, [0\ 1])$. Suppose that A has two columns γ_1 and γ_2. Since we require that A is simple, these two columns are distinct, and there is some row r on which they differ. On this row, the matrix $A|_{\{r\}}$ will either be $[0\ 1]$ or $[1\ 0]$. In either case, $[0\ 1] \prec A$. We conclude that A cannot have two columns, and forb$(m, [0\ 1]) = 1$. □

Example 2. *Prove that forb* $\left(m, \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}\right) = m + 1.$

Solution. We give an inductive proof. The base case, $m = 2$, is trivial.

Suppose that the result is true for $m-1$ and let $A \in$ Avoid $\left(m, \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}\right)$. Expanding along the top row, we write A as follows:

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ B & C & C & D \end{bmatrix},$$

where the matrices B and D contain no columns in common. We note that if $[0\ 1] \prec C$, then $\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \prec A$. Hence, using Example 1, C contains at most one column.

Now we remove, from A, the top row and one copy of C to obtain

$$A' = [B\ C\ D] \in \text{Avoid}\left(m-1, \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}\right).$$

By the induction hypothesis, $||A'|| \leq m$. As we have only removed $||C|| \leq 1$ column from A to produce A', we have that $||A|| \leq m+1$ as required.

$\square$

With a few examples under our belt, we will now define the most important matrices for our discussions. Define K_k to be the matrix consisting of all possible columns on k rows, called the *complete matrix* on k rows. For the submatrix of K_k consisting of all columns with column sum ℓ, we use the notation K_k^ℓ.

$$K_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad K_3^2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Notice that for $\ell = 2$ the matrix K_k^ℓ gives the vertex-edge incidence matrix for a complete graph. Because of this, K_3^2 is often called a *triangle*. Ryser [?] proved that $\text{forb}(m, K_3^2) = \binom{m}{2} + \binom{m}{1} + \binom{m}{0}$, with a matrix proof using the linear algebra ideas of null t-designs in the case $t = 2$.

We are now ready to state the "fundamental theorem" of forbidden configurations. This result was proved independently by Sauer [?], Perles and Shelah [?], Vapnik and Chervonenkis [?].

Theorem 1.3. *(Sauer [?], Perles and Shelah [?], Vapnik and Chervonenkis [?]) Let k be given. Then*

$$\mathit{forb}(m, K_k) = \binom{m}{k-1} + \binom{m}{k-2} + \cdots + \binom{m}{0}.$$

We give two linear algebraic proofs of this result. The first proof uses ideas of null t-designs [?] and is due to Frankl and Pach (somewhat after the fact proven by Anstee [?]). The second proof is a wonderful example of the polynomial method due to Smolensky [?] and appears in Section 3. In this thesis, the *polynomial method* employs multilinear polynomials in variables $x_1, x_2, \ldots, x_m$ that are mostly zero on certain vectors and nonzero on a few others. We apply a dimension argument that the number of linearly independent multilinear polynomials of degree at most t is at most $\binom{m}{t}+\binom{m}{t-1}+\cdots+\binom{m}{0}$.

There have been many, more general, versions of this result, that have been proved in a variety of ways. Our linear algebra techniques apply themselves nicely to one such generalization, which rears it's head in each section of our paper. Instead of avoiding a single configuration F, we will first restrict which set of rows S we are looking at, and then choose a configuration F_S to restrict on these rows. This idea appeared first in Alon [**?**], to the best of the Author's knowledge.

Definition 1.4. *Given a set S and a family $\mathcal{F}\subseteq 2^S$, we define the* downset *of $\mathcal{F}$ to be the collection $\mathcal{D}(\mathcal{F})$ of all sets that do not contain any element of $\mathcal{F}$.*

$$\mathcal{D}(\mathcal{F})=\{T\subseteq S : F\not\subseteq T \textit{ for any } F\in\mathcal{F}\}.$$

This downset is strict in the sense that $\mathcal{D}(\mathcal{F})$ does not contain any elements of $\mathcal{F}$. It may also be useful to refer to a non-strict downset, which we will identify unambiguously as $\mathcal{D}(\mathcal{F})\cup\mathcal{F}$.

We generalize Theorem **??** as follows.

Theorem 1.5. *(Alon [**?**]) Let A be a simple $\{0,1\}$-matrix, and let $[m]$ enumerate the rows of A. Let $\mathcal{S}\subset 2^{[m]}$ be a collection of subsets of the rows such that for each $S\in\mathcal{S}$, $K_{|S|}\nprec A|_S$. Then $||A||$ is bounded by $|\mathcal{D}(\mathcal{S})|$.*

Note that this is indeed a generalization of Theorem **??**. If we consider $\mathcal{S}=\binom{[m]}{k}$, then the condition $K_{|S|}\nprec A|_S$ for every $S\in\mathcal{S}$ is equivalent to $K_k\nprec A$. Hence, with this choice of $\mathcal{S}$, Theorem **??** and Theorem **??** have the same hypotheses. The conclusion is also the same. When $\mathcal{S}=\binom{[m]}{k}$, we have $\mathcal{D}(\mathcal{S})=\binom{[m]}{k-1}\cup\binom{[m]}{k-2}\cup\cdots\cup\binom{[m]}{0}$ so that $|\mathcal{D}(\mathcal{S})|=\binom{m}{k-1}+\binom{m}{k-2}+\cdots+\binom{m}{0}$.

Two proofs of Theorem **??** appear in Sections **??** and **??**. In Section **??** we see a linear algebraic proof involving null t-designs. And in Section **??**, the polynomial method is used to obtain a short proof of this result.

Section **??** is concerned with violations. Let A be an $m\times n$ simple $\{0,1\}$-matrix and t a constant. Given a k-set S of the rows, we say a vector $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_k)^T$ is in short supply if it appears at most t columns of $A|_S$. Row order matters here; we are not using α as a configuration. We say that we have a *violation* on S if there is a column γ of A with α appearing in $\gamma|_S$. We prove that if there are two vectors α and β that are in short supply on every $S\in\binom{[m]}{k}$, then we can remove $O(m^{k-1})$ columns from A to get rid of all the violations. This is a remarkable result because you may expect there to be as many as $t\binom{m}{k}$ violations to remove. The proof is also a nice application of the polynomial method.

We use the results of Section **??** to prove the main results of Section **??**. We find asymptotic bounds for two configurations, F_k, and $G_k(B)$. We show that $\mathrm{forb}(m,F_k(B))$

and $\text{forb}(m, G_k)$ are $\Theta(m^{k-1})$. These configurations are interesting because they are maximal configurations on k rows with $\text{forb}(m, F) = \Theta(m^{k-1})$. We follow the linear algebra proofs of Anstee, Fleming, Füredi and Sali [**?**], and Anstee and Fleming [**?**].

2 Null t-Designs

Given v, k, t, λ we say a multiset $\mathcal{B}$ is a t-design of multiplicity λ if each $B \in \mathcal{B}$ is a k-set of $[v]$ (i.e. $B \in \binom{[v]}{k}$) and for each $T \in \binom{[v]}{t}$, there are precisely λ sets B in $\mathcal{B}$ with $T \subset B$. We allow more than one copy of B in $\mathcal{B}$ so that trivially if we have a t-design for v, k, λ, then we have a t-design for $v, k, 2\lambda$ by taking the multiset consisting of two copies of $\mathcal{B}$. A *simple* design is one which has no repeated sets in $\mathcal{B}$ i.e. $\mathcal{B} \subseteq \binom{[v]}{k}$. The recent breakthrough results of Keevash [**?**] shows that simple t-designs exist subject to easy divisibility conditions. *Null t-designs* are introduced by Frankl and Pach [**?**] to consider a weighted set system (we allow sets to have both positive multiplicity and negative multiplicity) which is a t-design with multiplicity 0 when we count sets B in $\mathcal{B}$ with $T \subset B$ according to their weights/multiplicity. As well, they do not require the sets B to all be of the same size k. We follow their proof and null t-designs appear in the proof of Theorem 2.1 near the end of this section. In fact Ryser [**?**] considered the case $t = 2$.

Let $A = (a_{ij})$ be an $m \times n$ $\{0,1\}$-matrix. The rows of this matrix can be thought of as subsets of an n-element set, where an element is in the set if and only if the corresponding entry in the row vector is 1. Given a subset $S \subset [m]$, we form the row intersection vector $A(S) = (a(S)_j)_{j=1}^n$ where $a(S)_j = \prod_{i \in S} a_{ij}$. This definition $A(S)$ gives exactly the intersection of all the sets A associated to the row vectors in S.

To illustrate these definitions, consider

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

If $S = \{1, 2, 4\}$ then $A(S) = (1, 0, 1, 0, 1, 0)$ and $A|_S = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$. Notice that $A(S)$ contains a 1 in column α if and only if column α of $A|_S$ is entirely $1's$.

The following theorem relates the number of columns of A to the dimension of the vector space spanned by certain row intersection vectors.

Theorem 2.1. *[?] Let t be fixed, and let A be an $m \times n$ $\{0,1\}$-matrix with no configuration K_t. Then the dimension of the vector space spanned by all k-fold row intersection vectors of A for $k < t$, is equal to the number of distinct columns of A.*

Notice that, in the above theorem, we do not require our matrix A to be simple. We may take the matrix A to be simple without losing any generality. Indeed, if we have an $m \times n$ simple matrix A, consider a new matrix A' by appending an $n+1$st column identical to, for example, the 17th column of A. The vector space spanned by the row intersection vectors of A' will have the same dimension of that of A. All we have done is added an $n+1$ coordinate that is identical to the 17th.

Notice also that one direction of the above theorem is trivial. As the row intersection vectors are all of length equal to the number of columns of A, the dimension of the vector space they span is at most the number of columns of A.

We begin with three combinatorial lemmas. Note that in the following we do not require our matrices to be simple.

Definition 2.2. *Given an matrix A, define $\sigma_1(A)$ to be the counting function of the number of 1's in A.*

For example, for the given 1×3 matrix, $\sigma_1(\begin{pmatrix}1 & 0 & 1\end{pmatrix}) = 2$. Viewing row and column vectors $1 \times n$ and $m \times 1$ matrices respectively, we will use this notation for the number of ones in a vector as well.

The notation E_k and O_k will be reserved for the matrices on k rows containing the columns of even and odd column sums, respectively.

$$E_3 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \quad O_3 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

We use the notation $E_k^{\geq s}$ and $O_k^{\geq s}$ to denote the matrices on k rows consisting of all columns sums greater than or equal to s, with column sums even and odd, respectively.

Lemma 2.3. *Let A and B be non-empty k-rowed $\{0,1\}$-matrices with no columns in common and with*

$$\sigma_1(A(S)) = \sigma_1(B(S))$$

for all proper subsets $S \subsetneqq [k]$. Then there exists a $t > 0$ so that $A = t \cdot E_k$ and $B = t \cdot O_k$ or vice versa.

Proof. Without loss of generality, say that A contains a column $\mathbf{u}$ with a maximal number of 1's between all columns of A and B. We claim that this vector is a column of all 1's (i.e. a column of k 1's). Say that $\mathbf{u}$ has 1's on some set $S \subset [k]$ and 0's on the complement of S, and suppose to a contradiction that $[k] \setminus S \neq \emptyset$. Then we have by assumption $\sigma_1(B(S)) = \sigma_1(A(S)) > 0$. However, by the maximality of $\mathbf{u}$, B cannot contain a vector larger with more 1's than $\mathbf{u}$, and as A and B have no columns in common, we know $\mathbf{u}$ is not a column of B. Hence $\sigma_1(B(S)) = 0$, a contradiction. We can say without loss of generality that A contains a column of all 1s.

Say that A contains t columns consisting of k 1's. Then for every subset of rows S with $|S| = k-1$, we have $\sigma_1(A(S)) \geq t$. Hence for B as well we have $\sigma_1(B(S)) \geq t$. However, B contains no column consisting entirely of 1's, so we conclude B must exhibit every column with $k-1$ 1's. Indeed, each such column appears exactly t times in B. Furthermore, A cannot contain any of the columns with exactly $k-1$ 1's, since B contains them. This forms the base case for an induction argument.

We begin the induction step. Suppose that we have shown that A contains t copies of every column containing $k, k-2, k-4, \ldots, k-2l+2$ ones and B has been shown to contain every column consisting of $k-1, k-3, \ldots, k-2l+1$ ones. We now examine subsets of rows S with $|S| = k-2l$. Looking just at the columns we know to exist in A and B we have

$$\sigma_1(A(S)) \geq t\left(1 + \binom{2l}{2} + \binom{2l}{4} + \cdots + \binom{2l}{2l-2}\right)$$

and

$$\sigma_1(B(S)) \geq t\left(\binom{2l}{1} + \binom{2l}{3} + \cdots + \binom{2l}{2l-1}\right).$$

The only columns not counted in the above calculation are the columns $\mathbf{u}$ with $\sigma_1(\mathbf{u}) = k-2l$. The columns with a smaller number of 0's will always contribute a 0 to the row intersection $A(S)$ and $B(S)$. By assumption, we have $\sigma_1(A(S)) = \sigma_1(B(S))$, so we place t copies of all columns with column sum $k-2l$ in A. When we have done this, we calculate

$$\sigma_1(A(S)) = t\left(1 + \binom{2l}{2} + \binom{2l}{4} + \cdots + \binom{2l}{2l-2} + \binom{2l}{2l}\right)$$

and

$$\sigma_1(B(S)) = t\left(\binom{2l}{1} + \binom{2l}{3} + \cdots + \binom{2l}{2l-1}\right),$$

which are indeed equal because the sum of the even binomial coefficients equals the sum of the odd binomial coefficients.

This process continues, alternating between A and B, until we conclude that A contains t copies of all columns with column sum $k, k-2, \ldots, 2$ and B contains t copies of all columns with column sum $k-1, k-3, \ldots, 1$. (Here we have assumed that the parity of k is even. The case where the parity of k is odd is similar). It turns out we also place t copies of the column $\mathbf{0}$ with A. Since $\sigma_1(A(\emptyset)) = \sigma_1(B(\emptyset))$, both matrices A and B have the same number of columns. We note that

$$\sum_{i \text{ even}} \binom{k}{i} = 1 + \sum_{\substack{i \text{ even} \\ i \geq 2}} \binom{k}{i} = \sum_{i \text{ odd}} \binom{k}{i}$$

and so A, the matrix with even columns sums, will contain t copies of the 0 column.

We have showed that one of either A or B is $t \cdot E_k$, and the other is $t \cdot O_k$, as required. □

This proof was essentially done by Ryser [**?**] which led to his proof [**?**] that $\text{forb}(m, K_3^2) = \binom{m}{2} + \binom{m}{1} + \binom{m}{0}$. What happens if the matrix does not have exactly k rows? And what happens when we only require that the row intersections of A and B are equal only on specific subsets of the rows? The answer is a simple corollary to the previous lemma.

Lemma 2.4. *Let $\mathcal{S}$ be a family of subsets of the rows. Let A and B be non-empty $m \times n$ matrices with no columns in common and suppose $\sigma_1(A(S)) = \sigma_1(B(S))$ for every $S \in \mathcal{D}(\mathcal{S})$. Then there exists an $S \in \mathcal{S}$ so that $A|_S$ contains a copy of $O_{|S|}$ and $B|_S$ contains a copy of $E_{|S|}$ or vice versa.*

Proof. Let T be a non-empty minimal subset with $\sigma_1(A(T)) \neq \sigma_1(B(T))$. Note that, by minimality, $\sigma_1(A(S)) = \sigma_1(B(S))$ for all $S \subsetneqq T$.

Throw out all columns in common between $A|_T$ and $B|_T$ and appeal to Lemma **??** with $k = |T|$. We conclude that, when restricted to T, one of either A or B contains all even subsets and the other contains the odd subsets. Furthermore, as $\sigma_1(A(T)) \neq \sigma_1(B(T))$, we have $T \notin \mathcal{D}(\mathcal{S})$. Hence, there exists some $S \in \mathcal{S}$ with $S \subseteq T$. On this subset, $A|_S$ and $B|_S$ contain a copy of $O_{|S|}$ and $E_{|S|}$, as required. □

A different generalization of Lemma **??**, as proved in [**?**], is to place restrictions on the column sums of the matrices A and B. other words, we do not want our columns to be "small". The proof is no longer a direct corollary of Lemma **??**, but we follow the proof very closely. The only difference is that the induction step terminates at a different point.

Lemma 2.5. *Let s, t be given with $0 \leq s \leq t$. Let A and B be non-empty $\{0,1\}$-matrices with column sums at least s and no matching columns. If $\sigma_1(A(S)) = \sigma_1(B(S))$ for every $s \leq |S| < t$, then one of A or B has the configuration $O_t^{\geq s}$ and the other has the configuration $E_t^{\geq s}$.*

Proof. Without loss of generality, suppose that A contains a column $\mathbf{u}$ with a maximal number of 1's between A and B. We claim that $\sigma_1(\mathbf{u}) \geq t$. Indeed, suppose to a contradiction that $\sigma_1(\mathbf{u}) < t$, and let S be the set of rows on which the value of $\mathbf{u}$ is 1. Then, by assumption, $\sigma_1(A(S)) = \sigma_1(B(S))$, but as A and B have no columns in common and $\mathbf{u}$ was chosen to be maximal, we have a contradiction. It follows that $\sigma_1(\mathbf{u}) \geq t$. Letting S be the set of rows for which the value of $\mathbf{u}$ assumes a 1, we choose a minimal subset $T \subseteq S$ with the property that $\sigma_1(A(T)) \neq \sigma_1(B(T))$.

We now restrict our attention to $A|_T$ and $B|_T$. Remove all columns common to these two matrices, and all the new matrices on $|T|$ rows A' and B'. Note the equalities $\sigma_1(A'(S)) = \sigma_1(B'(S))$ for all $S \subsetneq T$ and $\sigma_1(A'(T)) \neq \sigma_1(B'(T))$. Observe that one of the matrices A' or B' must contain t columns consisting entirely of 1's. This forms the base case for an induction in the same manner as Lemma **??**.

For the induction step, suppose without loss of generality that A' contains exactly t copies of each column of column sum $|T|, |T|-2, \ldots, |T|-2\ell+2$ and B' contains exactly t copies of each column with sums $|T|-1, |T-3|, \ldots, |T|-2\ell+1$. We claim that exactly t copies of each column of column sum $|T|-2\ell$ are in A'. Indeed, let S be an arbitrary set with $|S| = |T| - 2\ell$. Granting our claim, we compute the values

$$\sigma_1(A'(S)) = \binom{2\ell}{2\ell} + \binom{2\ell}{2\ell-2} + \cdots + \binom{2\ell}{0}$$

and

$$\sigma_1(B'(S)) = \binom{2\ell}{2\ell-1} + \binom{2\ell}{2\ell-3} + \cdots + \binom{2\ell}{1}$$

which are indeed equal, as required. Placing the columns of column sum $|T| - 2\ell$ elsewhere would disrupt this equality, and columns of column sum less that $|T| - 2\ell$ add values of 0 to the $(|T| - 2\ell)$-fold row intersections.

This induction continues until we reach columns of column sum s. After this, there are both no columns of column sum less than s, and no requirement of equality $\sigma_1(A'(S)) = \sigma_1(B'(S))$ for sets $|S| < s$.

It follows that A' and B' contain copies of $O_{|T|}^{\leq s}$ and $E_{|T|}^{\leq s}$ in some order. Hence, the original matrices A and B contain these matrices as well. As $|T| \geq t$ we are done. $\square$

These generalizations are not necessary to prove Theorem **??**. In fact, only Lemma **??** is necessary, but the generalizations seen in Lemmas **??** and **??** give us means to generalize Theorem **??**, which will be explored later in the section. First, we prove Theorem **??**.

Proof of Theorem **??**. Without loss of generality, the columns of A are distinct. If they are not distinct, we can remove the repeated columns without affecting the linear dependence of the row intersection vectors.

Construct a matrix B whose rows are the row intersection vectors $A(S)$ where $S \in \mathcal{D}(\binom{[m]}{t-1})$. The dimension of the vector space spanned by $\{A(S) | S \in \mathcal{D}(\binom{[m]}{t-1})\}$

is at most n because it consists of vectors of length n. The interesting part of the theorem is that the dimension is exactly n.

For the sake of contradiction, suppose that the dimension is less than n. Then the columns of B also form a vector space with dimension less than n. Label the columns $B_1, B_2, \ldots, B_n$. Then there is a non-trivial relationship $x_1B_1 + \cdots + x_nB_n = 0$. Because all the vectors B_i contain only rational entries, the x_i can be chosen to be rational. We can obtain a solution where the x_i are integral by multiplying by a large enough integer. Form two matrices A^+ and A^- as follows. For each $i = 1, 2, \ldots, n$, if x_i is positive, column i of matrix A is placed x_i times in A^+. Instead, if x_i is negative, we place column i of matrix A in A^-, with multiplicity $-x_i$.

As an aside, notice that A with column multiplicities x_i for column i can be viewed as a null t-design as we introduced this at the beginning of the section.

Now A^+ and A^- satisfy the conditions of lemma **??**. Namely, A^+ and A^- are non-empty matrices on n rows and no columns in common. The family $\mathcal{S}$ here consists of all sets with t elements. Since $x_1B_1+\cdots+x_nB_n = 0$, and B is defined to be the matrix consisting of row intersection vectors of A, we have that $\sigma_1(A^+(S)) = \sigma_1(A^-(S))$ on $\mathcal{D}(\mathcal{S})$. We conclude that one of A^+ and A^- contains a copy of O_t and the other contains a copy of E_t. This shows that A contains a copy of K_t, a contradiction. Hence the vectors $A(S)$ are independent as claimed. □

As a simple corollary of this theorem, we give our first proof of Theorem **??**.

Proof of Theorem **??**. Suppose that $A \in \text{Avoid}(m, K_t)$. Then, by Theorem **??**, the number of distinct columns of A is equal to the dimension of the vector space spanned by all k-fold row intersection vectors, for $k < t$. As there are at most $\binom{m}{k-1} + \binom{m}{k-2} + \cdots + \binom{m}{0}$ row intersection vectors, the dimension of the vector space and hence the number of columns of A, is bounded by this number. □

An interesting corollary is that if $A \in \text{Avoid}(m, K_k)$ with $||A|| = \text{forb}(m, K_k)$, then the row intersection vectors $A(S)$ are all linearly independent. Finally, we apply this proof technique to prove Theorem **??**. First, we generalize Theorem **??** as follows. The proof follows much the same lines as the proof of Theorem **??** above.

Theorem 2.6. *Let $\mathcal{S}$ be a family of subsets of the rows, and let A be an $m \times n$ $\{0,1\}$-matrix so that for each $S \in \mathcal{S}$ there is no configuration $K_{|S|}$ in $A|_S$. Consider the vector space spanned by $A(D)$ for each $D \in \mathcal{D}(\mathcal{S})$. The dimension of this vector space is equal to the number of distinct columns of A.*

Proof. Without loss of generality, the columns of A are distinct. Construct a matrix B whose rows are the vectors $A(D)$, for each $D \in \mathcal{D}(\mathcal{S})$. The dimension of the vector space spanned by $\{A(D)|D \in \mathcal{D}(\mathcal{S})\}$ is at most n because the vectors are of length n. We will show that the dimension is exactly n.

Suppose, in order to reach a contradiction, that the dimension is strictly less than n. Then the columns of B will also form a vector space with dimension less than n. Label the columns $B_1, B_2, \ldots, B_n$. Because the dimension is less than n, there is a non-trivial relationship $x_1 B_1 + x_2 B_2 + \cdots + x_n B_n$, where the x_i can be assumed to be integers. Now, form two matrices A^+ and A^- as follows. For each $i = 1, 2, \ldots, n$, if x_i is positive, column i of matrix A is placed x_i times in A^+. On the other hand, if x_i is negative, we place column i of matrix A into A^-, with multiplicity $-x_i$.

Now A^+ and A^- satisfy the conditions of Lemma **??**. It follows that there is some $S \in \mathcal{S}$ so that A^+ contains a copy of $O_{|S|}$ and A^- contains a copy of $E_{|S|}$, or vice versa. In either event, the matrix A will contain as a configuration $K_{|S|} = [E_{|S|}\ O_{|S|}]$. As $S \in \mathcal{S}$, this yields the desired contradiction Hence, the dimension is exactly n, as required. □

All the hard work has been done; the proof of Theorem **??** is now easy.

Proof of Theorem **??**. Suppose that A is a simple $\{0,1\}$-matrix on m rows, and that $\mathcal{S}$ is a collection of subsets of the rows on which $K_{|S|} \nprec A|_S$ for every $S \in \mathcal{S}$. Then, by Theorem **??**, the number of distinct columns is equal to the dimension of the vector space spanned by $A(D)$, for $D \in \mathcal{D}(\mathcal{S})$. As there are only $|\mathcal{D}(\mathcal{S})|$ vectors, this serves as an upper bound for the number of columns in A. □

3 Smolensky's Simple Proof

Smolensky [**?**] gives an interesting and short proof of the result of Sauer, Perles and Shelah, Vapnis and Chervonenkis. The motivation for Smolensky's proof was because there was a strong interest in VC-dimension in machine learning. For our purposes, the proof is an elementary exhibition of the polynomial method. The goal is to associate the columns of a matrix A with linearly independent polynomials. Any bound on the dimension of the space of polynomials used thus also gives a bound on the number of columns of A. The idea is simple and the linear algebra is nice - the difficult part about a proof like this is coming up with what polynomials to use.

Recall that polynomials can be thought of both as functions and as objects in a vector space. Polynomials can have many variables. A polynomial is called *multi-linear* if it is linear in each of its variables.

Thinking of a polynomial in m variables as a function in $\mathbf{x} = (x_1, x_2, \ldots, x_m)^T$, we can define an *indicator function* to be a function f that has a non-zero value when it's input satisfies certain criteria, and evaluates to 0 otherwise.

Without further ado, here is the proof, due to Smolensky [**?**].

Proof of Theorem **??**. We assign an indicator polynomial to each column of A. If the i^{th} column of A is $\gamma_i = (a_{1i}, a_{2i}, \ldots, a_{mi})^T$, we define

$$f_i(\mathbf{x}) = \prod_{j=1}^{m}(x_j + a_{ji} - 1).$$

If $\mathbf{x}$ is taken to have $x_j \in \{0, 1\}$ we have

$$f_i(\mathbf{x}) = \begin{cases} (-1)^{\sigma_0(\gamma_i)} & \text{when } \mathbf{x} = \gamma_i \\ 0 & \text{otherwise} \end{cases}$$

All of these polynomials are linearly independent. Indeed, the points $\gamma_1, \gamma_2, \ldots, \gamma_n \in \mathbf{R}^m$ exhibit n points which evaluate to non-zero values on precisely one polynomial.

$$f_i(\gamma_j) = \begin{cases} \neq 0 & \text{when } i = j \\ 0 & \text{otherwise} \end{cases}$$

In order to bound the number of columns of A, we can instead place a bound on the possible number of polynomials f_i. We immediately note that each polynomial is multi-linear of degree m. Using just this fact, the number of such polynomials is bounded by 2^m, the dimension of this space of polynomials. However, we can improve this bound.

We use the fact that $K_k \not\prec A$. For each k-tuple of rows $S = \{j_1, j_2, \ldots, j_k\} \subset [m]$, there must be some $\{0,1\}$-vector $\alpha = \alpha_S = (\alpha_{j_1}, \alpha_{j_2}, \ldots, \alpha_{j_k})$ which does not appear in $A|_S$. We form the polynomials

$$g_S(\mathbf{x}) = \prod_{j \in S}(x_j + \alpha_j - 1).$$

As each column of A avoids α_S on S, we have $g_S(\gamma_i) = 0$ for all i. This gives us an identity

$$g_S(\gamma_i) = \prod_{j \in S}(a_{ji} + \alpha_j - 1) = \prod_{j \in S} a_{ji} + h_S(\gamma_i) = 0$$

where h_S is a polynomial of degree less than k. Upon rearranging, we get the following identity, which holds for all columns γ_i.

$$\prod_{j \in S} a_{ji} = -h_S(\gamma_i)$$

We will use these identities to reduce each f_i to polynomials of degree less than k. Suppose that the monomial $c_J \prod_{j \in J} x_j$ is a term in the polynomial f_i. If $S \subseteq J$, we can write

$$\prod_{j \in J} x_j = \prod_{j \in S} x_j \cdot \prod_{j \in J \setminus S} x_j = -h_S \cdot \prod_{j \in J \setminus S} x_j,$$

where the equality holds only when evaluated at columns γ_i of A. The benefit of doing this substitution is that it reduces the degree of the polynomial f_i. We repeat doing substitutions in this manner to f_i until we can no longer do any more substitutions. The result, after many iterations, will be a new polynomial f_i' which is still multi-linear, but has degree less than k and satisfies $f_i'(\gamma_j) = f_i(\gamma_j)$ for all columns γ_j. Hence, we can bound the number of polynomials f_i by the number of multi-linear polynomials of degree less that k. This number is $\binom{m}{k-1} + \binom{m}{k-2} + \cdots + \binom{m}{0}$ and we are done. $\square$

Amazingly, Smolensky's proof can be easily modified to give a proof of Theorem **??**.

Proof of Theorem **??**. We follow closely the previous proof. Let γ_i be the i-th column of A, and define the polynomial

$$f_i(\mathbf{x}) = \prod_{j=1}^{m}(x_j + a_{ji} - 1)$$

as before. Note that $f_i(\gamma_i) \neq 0$ and $f_i(\gamma_j) = 0$ for $j \neq i$, so that our polynomials are linearly independent.

For each $S = \{s_1, s_2, \ldots, s_{|S|}\} \in \mathcal{S}$, there is some column missing from $A|_S$. Let $\alpha_S = \{\alpha_{s_1}, \alpha_{s_2}, \ldots, \alpha_{s_{|S|}}\}$ be one such column. Define the polynomials

$$g_S(\mathbf{x}) = \prod_{j \in S}(x_j + \alpha_j - 1) = \prod_{j \in S} x_j + h_S(\mathbf{x}).$$

Notice that $g_S(\gamma_i) = 0$ for any column γ_i. Hence, we can use the reduction

$$\prod_{j \in S} x_j = -h_S(\mathbf{x})$$

to reduce the degree of any monomial in f_i, for each $i = 1, 2, \ldots n$. These reductions will not change the value of f_i when evaluated at a column of A. After making all possible reductions, call the new polynomials $f_i'(\mathbf{x})$. At the end of the reductions, we note that no monomial in any f_i will contain a product $\prod_{j \in S} x_j$ for any $S \in \mathcal{S}$.

We note that each f_i' satisfies

$$f_i'(\gamma_j) = \begin{cases} \neq 0 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

and hence $f_1', f_2', \ldots, f_n'$ are linearly independent. As every monomial in f_i is a constant times a product of terms $\prod_{j \in D} x_j$, where $D \in \mathcal{D}(\mathcal{S})$, the number of such polynomials is bounded above by $|\mathcal{D}(\mathcal{S})|$. This, in turn, gives the desired bound on the number of columns in A. $\square$

4 Violations

Given an $m \times n$ simple $\{0,1\}$-matrix A, and a k-element set S of the rows and a fixed integer t, we say that a k-element column $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)^T$ is in *short supply* in A on S if the submatrix α occurs in $A|_S$ less than t times. When a column α is in short supply on S, we say that a column $\gamma \in A$ is a *violation* if $\gamma|_S = \alpha$.

If F is a $k \times \ell$ configuration and F is not a configuration of A, then for each $S \in \binom{[m]}{k}$, some columns must be either absent or in short supply. Hence, if we can remove all the columns with violations, then on all rows S, we have $K_{|S|} \not\prec A|_S$, and we can apply Theorem **??**. This idea of removing all columns with violations is central to the results in this section and the next.

The following result will be used in Theorems **??** and **??**.

Theorem 4.1. *[**?**] Let A be an $m \times n$ simple matrix, and let $\mathcal{S} \subseteq \binom{[m]}{k}$ be a family of k-sets of the rows for which two columns α_S and β_S occur less than t times on $A|_S$. Then we can remove $O(m^{k-1})$ columns of A so that there are no violations α_S or β_S for every $S \in \mathcal{S}$.*

Note that the matrix A in the above theorem has as many as $2t\binom{m}{k}$ violations. Remarkably, all of these violations can be eliminated by removing only $O(m^{k-1})$ columns of A.

We begin with two interesting proofs of specific cases of this theorem, observed by Anstee and Sali. These proofs require elementary graph theory, and were the inspiration for Theorem **??**.

Theorem 4.2. *If A is a simple $m \times n$ matrix, and $\mathcal{S} \subseteq \binom{[m]}{2}$ a family of subsets of the rows for which $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ occur at most t times on $A|_S$ for every $S \in \mathcal{S}$. Then we can remove $2tm$ columns of A so that there are no longer any violations on $\mathcal{S}$.*

Proof. Begin by creating a graph G with vertex set indexing the rows of A, and edge set $\mathcal{S}$.

Suppose G contains an odd cycle and let $\alpha = (\alpha_1, \ldots, \alpha_m)^T$ be a column of A. Colour a vertex $i \in G$ with colour α_i. Since G contains an odd cycle of length $2\ell + 1$, the chromatic number $\chi(G)$ is at least 3. As we have coloured the vertices with only two colours, there will always be two vertices i, j, adjacent in the cycle, with the same colour. In particular, this corresponds to every column of A having a violation on some pair of rows i and j which are an edge of the cycle. As there are at most $2t$ violations of an edge and $2\ell + 1$ edges in the cycle where $2\ell + 1 \leq m$, there are most $2tm$ columns of A and we are done.

Suppose now that G contains no odd cycle, hence G is bipartite. Let T be a maximal spanning forest of G. Note that the number of edges in T is $m - c$ where c

is the number of connected components of G. Form a new matrix A' by removing all columns that have a violation on rows i, j, where $\{i, j\} \in T$. We will have removed at most $2t(m-c) < 2tm$ columns. We claim that A' contains no violations for any $\{i, j\} \in \mathcal{S}$. Suppose to a contradiction that A' contains a violation on rows i, j and column α. Colour vertex k of G by the values α_k of α. By construction of G, ij is an edge in G. Also, by construction of A', $ij \notin T$. A colouring of G that has no violations on edges of T must be a standard two colouring of each component of T and hence a two colouring of G as well. Then there is no violation on ij. Hence A' contains no violations. □

Theorem 4.3. *If A is a simple $m \times n$ matrix, and $\mathcal{S} \subseteq \binom{[m]}{2}$ a family of subsets of the rows for which $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ occur at most t times on $A|_S$ for every $S \in \mathcal{S}$. Then we can remove $2t(m-1)$ columns of A so that there are no longer any violations on $\mathcal{S}$.*

Proof. Create a graph G where the vertices of the graph index the m rows of A, and the edge set is given by $\mathcal{S}$. Let T be a maximal spanning forest in G. Remove columns from A so that there are no longer any violations on T. Note there are at most $m-1$ edges in T. Thus we remove at most $2t(m-1)$ columns from A. We now claim that there are no more violations on any two rows of A. Suppose to a contradiction that there is a violation on rows i, j. As we have removed all violations from T, ij is not an edge in T. Also, from how we constructed G, ij cannot be an edge between the connected components of G. Hence, $T \cup ij$ contains a cycle C, and there is a column α on which $\alpha_i = 0$ and $\alpha_j = 1$ (or vice versa). Now $C \setminus ij$ is a path from i to j in T, with $\alpha_i \neq \alpha_j$. There must be two adjacent vertices u, v in this path with $\alpha_u \neq \alpha_v$. Hence column α is a violation on an edge of T, which contradicts that we have removed all violations from T. The result follows. □

We now prove the main result of the section, Theorem **??**. In the two preceding arguments, we chose columns to delete by choosing a spanning tree of a graph. In the general case below, we instead greedily choose columns to delete, and show using indicator polynomials that choosing $O(m^{k-1})$ columns is enough. This is a fundamentally different proof.

Proof of Theorem **??**. We greedily choose k-sets of rows $S_1, S_2, \ldots, S_r$ and columns $\gamma_1, \gamma_2, \ldots, \gamma_r$ so that there is a violation α_{S_i} or β_{S_i} on γ_i, but no violation α_{S_j} or β_{S_j} on γ_i for any $j < i$.

We continue greedily choosing sets and columns until we cannot find any more. Thus, we may assume there is no column γ and k-set of rows S for which there is a violation α_S or β_S on γ for which γ has no violation α_{S_i} or β_{S_i} for $i \leq r$. If this was not the case we could set $\gamma = \gamma_{r+1}$ and $S = S_{r+1}$ and continue our greedy choices of k-sets of rows.

Now form indicator polynomials

$$p_i^\alpha = \begin{cases} 1 & \text{when violation } \alpha_i \text{ on } S_i \\ 0 & \text{otherwise} \end{cases}$$

$$p_i^\beta = \begin{cases} 1 & \text{when violation } \beta_i \text{ on } S_i \\ 0 & \text{otherwise} \end{cases}$$

Note that as each S_i is a k-set of rows, these polynomials can be made to be monic of degree k. Define

$$p_i = p_i^\alpha - p_i^\beta = \begin{cases} \pm 1 & \text{when violation } \alpha_i \text{ or } \beta_i \text{ on } S_i \\ 0 & \text{otherwise} \end{cases}$$

The polynomials p_i have degree at most $k-1$. We claim that all the polynomials p_i are linearly independent. Form an evaluation matrix E where the ij^{th} entry is $p_i(\gamma_j)$. The matrix will have entries ± 1 along the main diagonal as there is some violation on column γ_i in rows S_i. Also, as there is no violation on column γ_i in rows S_j for $j < i$, the upper triangle will only contain values of 0. Hence the evaluation matrix E is lower triangular, and the polynomials p_i are linearly independent. The dimension of the vector space of linearly independent polynomials on m variables of degree less than k is $\binom{m}{k-1} + \binom{m}{k-2} + \cdots + \binom{m}{1} + \binom{m}{0} = O(m^{k-1})$.

To recapitulate, we greedily chose r k-sets S_i and r columns γ_i. The columns γ_i had violations on S_i but not on S_j for any $j < i$. We associated to each i an indicator polynomial p_i which was of degree at most $k-1$, and observed that these polynomials were linearly independent. Hence r is at most $\binom{m}{k-1} + \binom{m}{k-2} + \cdots + \binom{m}{1} + \binom{m}{0}$.

To finish the proof, we remove the $2t$ violations from each of the r chosen k-sets S_i. As t is fixed, and r is $O(m^{k-1})$, we have removed $O(m^{k-1})$ columns. By doing this, we have removed not just the violations from the r rows, but all violations from all $S \in \mathcal{S}$. Indeed, if there were another violation, we could have chosen it next in our greedy process. As the greedy process terminated, there are no more violations. □

We can generalize Theorem **??** as follows, to obtain a new theorem. Given a family of subsets $\mathcal{S} \subseteq 2^{[m]}$, define

$$d(\mathcal{S}) = \{T \in \mathcal{D}(\mathcal{S}) \,:\, \text{there exists } S \in \mathcal{S} \text{ with } T \subset S\}.$$

This definition includes all sets that are proper subsets of some $S \in \mathcal{S}$, and gives us the correct bound for the following theorem.

Theorem 4.4. *Let A be an $m \times n$ simple matrix, and let $\mathcal{S} \subseteq 2^{[m]}$ be a family of subsets of the rows for which two columns α_S and β_S occur at most t times on $A|_S$. Then we can remove $2t|d(\mathcal{S})|$ columns of A so that there are no violations α_S or β_S for every $S \in \mathcal{S}$.*

Proof. We proceed similarly to the proof of Theorem **??**. Greedily choose sets of rows $S_1, S_2, \ldots, S_r$ and columns $\gamma_1, \gamma_2, \ldots, \gamma_r$ until you can choose no more satisfying the following. For each i, $S_i \in \mathcal{S}$, and there is a violation on column γ_i in rows S_i, but not on rows S_j for $j < i$.

As before, for each i we form indicator polynomials p_i^α and p_i^β of degree $|S_i|$. The polynomial p_i^α is multi-linear, and $p_i^\alpha(\mathbf{x}) \neq 0$ only when there is a violation α_{S_i} on rows S_i of $\mathbf{x}$. Similarly, p_i^β is multi-linear and $p_i^\beta(\mathbf{x}) \neq 0$ only when there is a violation β_{S_i} on rows S_i of $\mathbf{x}$. Now form the polynomial $p_i = p_i^\alpha - p_i^\beta$. These polynomials have degree less than $|S_i|$, are multi-linear, and only contains monomials containing terms from $d(\mathcal{S})$. When we evaluate each of these polynomials p_i at $\gamma_1, \gamma_2, \ldots, \gamma_r$ gives an upper triangular matrix with non-zero diagonal entries. Hence, $p_1, p_2, \ldots, p_r$ are linearly independent.

We have associated to each chosen set of rows $S_1, S_2, \ldots, S_r$ a polynomial whose terms are in $d(\mathcal{S})$. As the polynomials are independent, it follows that $r \leq |d(\mathcal{S})|$. We delete at most $2t$ violations from each of the chosen sets, after which there are no more violations. □

5 Boundary between $O(m^{k-1})$ and $\Omega(m^k)$

It was conjectured by Anstee and Sali [**?**] that the maximal configurations on k rows with $\text{forb}(m, F) = \Theta(m^{k-1})$ are the configurations $G_k(B)$ and F_k given below. The bound for $G_k(B)$ was proven using induction by Anstee and Fleming [**?**] and using linear algebra/polynomial method by Anstee, Fleming, Füredi and Sali [**?**] while the bound for F_k was proven by Anstee and Fleming using linear algebra and the polynomial method [**?**].

The first k-rowed configuration $G_k(B)$ is defined as

$$G_k(B) = [K_k | t \cdot [K_k \setminus B]]$$

where B is a $k \times (k+1)$ matrix consisting of one column of each column sum. To be explicit, when $k = 3$, there are only two choices up to row and column permutations for G_k using the following two choices for B:

$$B \in \left\{ \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \right\}.$$

The second k-rowed maximal matrix is

$$F_k = [0 \,|\, t \cdot D_{12}]$$

where D_{12} is the matrix consisting of all non-zero columns γ such that $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] \neq \gamma|_{\{12\}}$ As an example, D_{12} is shown below for $k = 3$ rows.

$$D_{12} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

We begin by proving that the first configuration, $G_k(B)$, indeed has $\text{forb}(m, G_k(B)) = \Theta(m^{k-1})$. The proof is an easy corollary of Theorem **??**.

Theorem 5.1. *Let B be a $k \times (k+1)$ matrix consisting of one column of each column sum, and let $G = [K_k | t \cdot [K_k \setminus B]]$. Then $forb(m, G)$ is $\Theta(m^{k-1})$.*

Proof. Let $A \in \text{Avoid}(G)$. We begin by observing what occurs on each k-set of rows S_i of A. As $A|_{S_i}$ avoids G, there are two possibilities. First is that there is a column missing from $A|_{S_i}$. In this case, $K_k \nprec A|_{S_i}$ so $G \nprec A|_{S_i}$. The second possibility is that there are at least two columns in short supply. To see why it is not enough for one column to be in short supply, let α be the only column in short supply on $A|_{S_i}$. Rearrange the rows of G so that α is one of the columns of B. As all other rows are in long supply, this rearrangement of G is a configuration of $A|_{S_i}$. Hence it does not suffice to have only one column in short supply.

Let $\mathcal{S}$ be the set of k-sets of row S on which there are two columns α_S, β_S in short supply. In short supply. Apply Theorem **??**. We can remove $O(m^{k-1})$ columns to remove all violations from $\mathcal{S}$. Now every k-set of rows has a column missing, so by Theorem **??**, only $O(m^{k-1})$ columns remain. Hence $\text{forb}(m, G)$ is $O(m^{k-1})$.

A simple construction works to show it is possible to construct a matrix with $\Omega(m^{k-1})$ columns avoiding G - simply take all columns with column sum less than k. $\square$

In the same vein as other generalizations in this paper, we can prove a version of the result where we first choose sets S of columns, then choose matrices $G_{|S|}(B)$ to restrict on those rows. Notice that our choice of B can be different for each set $\mathcal{S}$, making this genuinely different from Theorem **??**, even when we take $\mathcal{S} = \binom{[m]}{k}$. This is a new result.

Theorem 5.2. *Let A be a simple $\{0,1\}$-matrix and $\mathcal{S}$ a collection of subsets of the rows. For each $S \in \mathcal{S}$, let B_S be a $|S| \times (|S|+1)$ matrix with one column of each column sum, and define $G_S = [K_{|S|} | t \cdot [K_{|S|} \setminus B_S]]$. If $G_S \nprec A|_S$ for every $S \in \mathcal{S}$, then $||A||$ is $\Theta(|\mathcal{D}(\mathcal{S})|)$.*

Proof. Let $S \in \mathcal{S}$. As $G_S \not\prec A|_S$, there is either a column missing from $A|_S$, or there are two columns α_S and β_S in short supply on S. Let $\mathcal{T}$ denote the subset of $\mathcal{S}$ on which there are two columns in short supply. Apply Theorem **??** to this set, so we can remove $O(|d(\mathcal{T})|)$ columns so that, on every $T \in \mathcal{T}$, there is some column missing. Hence, on every $S \in \mathcal{S}$, $K_{|S|} \not\prec A|_S$. By Theorem **??**, there are at most $|\mathcal{D}(\mathcal{S})|$ columns remaining. It follows that our original matrix A had at most $O(|\mathcal{D}(\mathcal{S})|)$ columns.

To exhibit a matrix A of this type with $\Omega(|\mathcal{D}(\mathcal{S})|)$ columns, simply take the matrix consisting of all columns corresponding to sets in $\mathcal{D}(\mathcal{S})$. □

We now tackle the trickier case. We follow the proof given by Anstee and Flemming [**?**], and we hope to have made the proof easier to digest with additional examples and explanations.

Theorem 5.3. *Let $F_k = [\mathbf{0}|t \cdot D_{12}]$. Then $forb(m, F_k)$ is $\Theta(m^{k-1})$.*

Theorem **??** was an easy application of Theorem **??**, since when a matrix avoids $G_k(B)$, then on every k-set of rows there is either a column missing or two columns in short supply . However, when a matrix A avoids F_k, it is also possible that on some k-sets of rows there is just one columns in short supply. In general, it can be quite complicated to describe exactly what is missing on a k-set of rows. The following proposition is sufficient for our proof.

Proposition 5.4. *Let A be a matrix with no configuration F_k and let S be a k-set of rows of A. Then at least one of the following cases occurs on S.*

a) The column of 0's is absent on $A|_S$.

b) There is a column $(0, 0, \dots, 0, 1)^T$ in short supply on S.

c) There are two or more columns of column sum at least 2 in short supply on S.

Furthermore, given any pair of rows $i, j \in S$ there is a column in short supply (or missing) on S which is either 0 on row i or 0 on row j.

Proof. If there is no column $\mathbf{0}$ contained in $A|_S$ then $F_k \not\prec A|_S$. Similarly, as $K_k^1 \prec F_k$, if some column $(0, 0, \dots, 0, 1)^T$ is in short supply on S, then $F_k \not\prec A|_S$. Hence, if either (a) or (b) holds, then A has no copy F_k. Now suppose that $A \in \text{Avoid}(F_k)$ and neither (a) nor (b) holds. We claim that in this case (c) holds. Indeed, if there is only one column α of column sum at least 2 in short supply, then we can permute the rows of A so that the column in short supply can be written $\alpha = (1, 1, \alpha_3, \dots, \alpha_k)$. With this ordering, there is no column of $D_{1,2}$ in short supply and hence $F_k \prec A|_S$. Hence one of (a), (b), and (c) are necessary for A to avoid F_k on S.

To prove the second claim, suppose there is a choice of rows $i, j \in S$ so that for each choice of column α in short supply, we have $\alpha|_{\{i,j\}} = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$. Permuting these two rows to the top, we see that no column of $D_{1,2}$ is in short supply and the column $\mathbf{0}$ is not absent. Hence $F_k \prec A|_S$ in this case, a contradiction. $\square$

In general, there will be some collection of sets of rows $\mathcal{T} \subset \binom{[m]}{k}$ with two columns in short supply. We will use Theorem **??** on $\mathcal{T}$ to remove all violations from these sets of rows. However, Proposition **??** tells us that there may be some collection $\mathcal{S}$ of k-sets of rows where there is only one column in short supply. For this collection $\mathcal{S}$, more work will have to be done in order to remove all violations from these rows. The next few propositions will detail how we can remove $O(m^{k-1})$ columns from A to remove all the violations in $\mathcal{S}$.

Observe that whenever there is only one column in short supply on a k-set S, by Proposition **??** that column is necessarily $(0, 0, \ldots, 0, 1)^T$. Now, we introduce notation which will be well-used in the coming discussions. For a $(k-1)$-set of rows $G = \{i_1, i_2, \ldots, i_{k-1}\}$ we say we have the *implication* $G \to i_k$ if $G \cup \{i_k\} \in \mathcal{S}$ and, with the given order, the column $(0, 0, \ldots, 0, 1)^T$ is in short supply on $G \cup \{i_k\}$. Hence, if we have the implication $G \to i_k$, then whenever a column $\gamma|_G = \mathbf{0}$, we almost always have $\gamma|_{i_k} = 0$ as well, with at most t exceptions. We extend the definition to include the vacuous implications $G \to i_1, G \to i_2, \ldots, G \to i_{k-1}$.

We are able to make the idea of an implication into a more symmetric and useful idea. Define a digraph $D(A)$ with vertex set $\binom{[m]}{k-1}$ and edges

$$G \to H \text{ if and only if } G \to h \text{ for all } h \in H. \tag{1}$$

For a given vertex G, we write $C(G)$ for the strongly connected component containing a set $G \in \binom{[m]}{k-1}$. Recall the following result from graph theory.

Proposition 5.5. *It is always possible to partition the vertices of a directed graph $G = (N, A)$ so that each partition $P_1, P_2, \ldots, P_r$ is strongly connected. Moreover, it is possible to order the partitions so that every edge $u \to v$ between partitions goes from a lower numbered partition to a higher numbered partition.*

Proof. First, note that the property "there exist directed paths from u to v and from v to u" forms an equivalence relation on vertices $u, v \in N$. Hence we can partition the vertices into what are called strongly connected components on vertex sets $P_1, P_2, \ldots$. Next, define a relation between partitions $P_i \to P_j$ if there exists an arc originating in P_i and terminating in P_j. Observe that this relation is acyclic, so there is an ordering of the partitions/strongly connected components as required. $\square$

The ordering of strongly connected components is often called the *topological ordering* or the *acyclic ordering* of strongly connected components. As the vertices of our digraph are $(k-1)$-sets of rows of A, we define the *support* of a strongly connected component $C \subseteq D(A)$ to be the set of all rows that are contained in some vertex $G \in C$. We denote the support of C as

$$\mathrm{supp}(C) = \cup_{G \in C} G.$$

We define an implication $G \to i$ to be *inner* if $i \in \mathrm{supp}(C(G))$ and we define an implication to be *outer* otherwise. Similarly, we say that a violation $(0, 0, \ldots, 0, 1)^T$ on rows $(i_1, i_2, \ldots, i_k)$ is *inner* (*outer*) if $\{i_1, i_2, \ldots, i_{k-1}\} \to i_k$ is inner (outer). In the coming proof, we will deal with the inner and outer violations separately.

These definitions will hopefully be made crystal clear with an example. Suppose that $A \in \mathrm{Avoid}(F_3)$ is a matrix on $m = 6$ rows. As A avoids F_3, there will be some columns in short supply on every 3-set of rows. We are concerned with the 3-sets of rows where there is only one column in short supply. Suppose that this family is

$$\mathcal{S} = \big\{\{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{1,3,6\}, \{1,4,6\}, \{2,3,4\}\big\}.$$

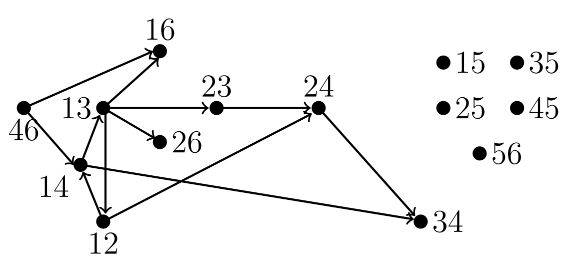
On these sets of rows, suppose the following columns are in short supply.

$$\begin{array}{c}1\\2\\3\\4\\5\\6\end{array}\begin{bmatrix}0\\1\\0\\-\\-\\-\end{bmatrix} \quad \begin{array}{c}1\\2\\3\\4\\5\\6\end{array}\begin{bmatrix}0\\0\\-\\1\\-\\-\end{bmatrix} \quad \begin{array}{c}1\\2\\3\\4\\5\\6\end{array}\begin{bmatrix}0\\-\\1\\0\\-\\-\end{bmatrix} \quad \begin{array}{c}1\\2\\3\\4\\5\\6\end{array}\begin{bmatrix}0\\-\\0\\-\\-\\1\end{bmatrix} \quad \begin{array}{c}1\\2\\3\\4\\5\\6\end{array}\begin{bmatrix}1\\-\\-\\0\\-\\0\end{bmatrix} \quad \begin{array}{c}1\\2\\3\\4\\5\\6\end{array}\begin{bmatrix}-\\0\\0\\1\\-\\-\end{bmatrix}$$

These correspond to the implications

$$\{1,3\} \to 2, \quad \{1,2\} \to 4, \quad \{1,4\} \to 3, \quad \{1,3\} \to 6, \quad \{4,6\} \to 1, \quad \{2,3\} \to 4.$$

We form the digraph $D(A)$. Note some isolated vertices.



Most arcs in the digraph feature trivial implications. For example, the implication $\{1,3\} \to \{1,6\}$ contains the implication $\{1,3\} \to 6$ and a trivial implication

$\{1,3\} \to 1$. The only arc of $D(A)$ which arises from only non-trivial implications is $\{1,3\} \to \{2,6\}$.

There is only one non-trivial strongly connected component of $D(A)$, consisting of the vertices $\{1,2\},\{1,3\},\{1,4\}$. The support of that component is $\{1,2,3,4\}$. Hence the following implications are inner:

$$\{1,3\} \to 2, \quad \{1,2\} \to 4, \quad \{1,4\} \to 3,$$

while the remaining implications are outer:

$$\{4,6\} \to 1, \quad \{1,3\} \to 6, \quad \{2,3\} \to 4.$$

Lemma 5.6. *Given a matrix A avoiding F_k, form the digraph $D(A) = (\binom{[m]}{k-1}, E)$ as above. We can remove $4t(k-1)\binom{m}{k-1}$ columns from A so that the resulting matrix contains no inner violations.*

Proof. Select a minimal subset of edges $E' \subseteq E$ so that the digraph $(\binom{[m]}{k-1}, E')$ has the same strongly connected components as $D(A)$. At most $2\binom{m}{k-1}$ arcs are needed. For each directed edge $G \to H \in E'$, take the $k-1$ implications $G \to h$, $h \in H$. Delete all columns with violations of the selected implications. As there are at most t violations for any selected implication, we have removed at most $4t(k-1)\binom{m}{k-1}$ columns. It is left to show that no inner violations remain.

Let $G \to h$ be inner, and take $H \ni h$ so that G and H are in the same strongly connected component. Choose a path $G \to G_1 \to \cdots \to G_s \to H$, where each directed edge is in E'. As there are no violations on any of the selected arcs, whenever a column $\gamma|_G = \mathbf{0}$, we inductively find that $\gamma_{G_i} = \mathbf{0}$ and $\gamma_H = \mathbf{0}$ as well. Whence, there are no more inner violations $G \to h$. □

Hence, we can remove all inner violations by removing only $O(m^{k-1})$ columns from A. Our goal is to do the same with the outer violations. To do so, we will make a reduction from degree $k+1$ polynomials to degree $k-1$ polynomials. Note the difference between Theorem **??**, which only results in a one degree reduction from k to $k-1$. The following lemma is central to the argument that follows.

We use the notation from Anstee and Fleming [**?**]. Given a set of rows S and a column α, define the slightly different indicator polynomial for the column restricted to S as follows.

$$f_{S,\alpha} = \prod_{r \in S} (x_r - \alpha_r) \tag{2}$$

Note that $x_r - \alpha_r \neq 0$ if $x_r \neq \alpha_r$, i.e. x_r is the (0,1)-complement of α_r. Thus this indicator polynomial has the property that $f_{S,\alpha}(\mathbf{x}) = 0$ for all $\{0,1\}$-vectors $\mathbf{x}$ except those with $\mathbf{x}|_S = \alpha^c$ where α^c denotes the $(0,1)$-complement of α. Using this alternate form of the indicator polynomials results in the following result for which a reduction from degree $k+1$ to degree $k-1$ is easier to deduce.

Lemma 5.7. *Let $S \subset \binom{[m]}{k+1}$ be a collection of $k+1$-sets and let $M = M_S$ be a matrix consisting of the columns in short supply on S. Assume that $\varepsilon = (\varepsilon_\alpha)$ satisfies $\sum_\alpha \varepsilon_\alpha = 0$ and $M\varepsilon = \mathbf{0}$. Use the new indicator polynomials $f_{S,\alpha}$ of (??). Define a new indicator polynomial for the columns α for which $\varepsilon_\alpha \neq 0$:*

$$p_S(\mathbf{x}) = \sum_\alpha \varepsilon_\alpha f_{S,\alpha}(\mathbf{x}). \tag{3}$$

Then $p_S(\mathbf{x})$ has degree $k-1$.

Proof. Expand the products and sum in the definition of p_S. Notice that the coefficient of the degree $k+1$ term is $\sum_\alpha \varepsilon_\alpha$, and the coefficient of the degree k monomial $\prod_{s \in S \setminus r} x_s$ is $-\sum_\alpha \alpha_r \varepsilon_\alpha$. Hence, the conditions $\sum_\alpha \varepsilon_\alpha = 0$ and $M\varepsilon = \mathbf{0}$ say exactly that these coefficients are 0. □

We demonstrate with an example how this lemma will be used in the general proof. Assume that $k = 4$ and there is a set $G = \{i_1, i_2, i_3\}$ with outer implications $G \to i_4$ and $G \to i_5$. Hence, the following three vectors are in short supply on $S = \{i_1, i_2, i_3, i_4, i_5\}$:

$$\begin{array}{c} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{array} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{array}{c} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{array} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad \begin{array}{c} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{array} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

There are other vectors in short supply on S as well. Indeed, for any $i = 1, 2, 3$ we can examine the set $S_i = S \setminus i$, which is a k-set of rows. Hence, by Proposition **??** there is some column $(a, b, c, d)^T$ in short supply on S_i, and furthermore, we may assume that one of c or d is equal to 0. So for $i = 3$, we may have the following column in short supply:

$$y_3 = \begin{array}{c} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{array} \begin{bmatrix} 1 \\ 1 \\ \\ 1 \\ 0 \end{bmatrix}.$$

We can fill in the blank entry with either a 0 or a 1, and both vectors will be in short supply on S to obtain vectors $y_3^0 = (1, 1, 0, 1, 0)^T$ and $y_3^1 = (1, 1, 1, 1, 0)^T$ respectively.

$$y_3^0 = \begin{array}{c} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{array} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad y_3^1 = \begin{array}{c} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{array} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

Note that $y_3^1 - y_3^0 = (0,0,1,0,0)^T$.

We can similarly define y_1, y_1^0, y_1^1, y_2, y_2^0, and y_2^1, and extend this definition to $y_4^0 = (0,0,0,0,1)^T$, $y_4^1 = y_5^1 = (0,0,0,1,1)^T$, and $y_5^0 = (0,0,0,1,0)^T$. Notice that in general, $y_i^1 - y_i^0 = e_i$ where e_i is the standard basis vector. Thus we have the relationship

$$y_3^0 = (y_1^1 - y_1^0) + (y_2^1 - y_2^0) + (y_4^1 - y_4^0) + 0(y_5^1 - y_5^0). \tag{4}$$

This is not a suitable relationship for Lemma **??** because the sum of the coefficients are not zero. We consider the additional relationship,

$$y_5^0 + y_4^0 - y_4^1 = \mathbf{0}.$$

The sum of coefficients in this equation is 1. In principle, we can multiply this relationship by any constant in order to make the coefficients sum to any number. For our purposes, the sum of coefficients in (**??**) is always 1. Hence, we find the following relationship satisfies the conditions of Lemma **??**.

$$y_3^0 - (y_1^1 - y_1^0) - (y_2^1 - y_2^0) - (y_4^1 - y_4^0) - (y_5^0 + y_4^0 - y_4^1) = \mathbf{0}.$$

This yields coefficients ϵ_α for the various columns on 5 rows and yields a degree 3 indicator polynomial using (**??**). In our lemma below, we find a degree $k-1$ polynomial which will be of use to us in our proofs.

Lemma 5.8. *Given a matrix A avoiding F_k, form the digraph $D(A) = (\binom{[m]}{k-1}, E)$ as described above. Suppose further that A contains no inner violations. Then we can remove $O(m^{k-1})$ columns from A so that the resulting matrix contains no outer violations.*

Proof. We begin by processing the outer implications. Let $\mathcal{O}$ denote the set of all outer implications $G \to i$ where of course $G \in \binom{[m]}{k-1}$. We process the vertices of $D(A)$, namely $(k-1)$-sets $\binom{[m]}{k-1}$, in an order which respects the topological ordering of $D(A)$. Namely, if $G, H \in \binom{[m]}{k-1}$ and G belongs to a strongly connected component which is ordered before the strongly connected component containing H, then we process G before H. When we process G we consider outer implications $G \to i$ in any order, and delete an implication if, whenever it is violated, some remaining implication is also violated. Repeat until all elements of $\binom{m}{k-1}$ have been processed. Call the remaining outer implication which have not been deleted $\mathcal{O}'$.

Claim 1. If $G \to x$ and $G \to y$ are in $\mathcal{O}'$ then there is no implication $G \setminus i \cup x \to y$ in the original digraph $D(A)$ for any $i \in G$.

To prove the claim, let $G \to x \in \mathcal{O}'$ and $G \to y \in \mathcal{O}'$ be outer implications. Let $H = G \setminus i \cup x$, and suppose that there is an implication $H \to y$. It suffices to show $G \to y \notin \mathcal{O}'$.

$$
G\left\{\begin{array}{c} i \\ \\ \\ \\ x \end{array}\right.
\begin{array}{c} \\ \\ \\ \\ \\ y \end{array}
\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ - \\ 1 \end{bmatrix}
\left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\}H
$$

Recall that $G \to j$ is a trivial implication for all $j \in G$, and note that $G \to H$ is an arc in $D(A)$ because $G \to h$ for all $h \in H$. As $G \to x$ is outer, G and H belong to different strongly connected components and it follows that G is processed before H in our formation of $\mathcal{O}'$. Now notice that a violation $G \to y$ forces a violation of either $G \to x \in \mathcal{O}'$ or the later outer implication $H \to y$, depending on the value of the column at x being 1 or 0. In either case, we contradict that $G \to y$ has been chosen for inclusion in $\mathcal{O}'$. Thus we cannot have the implication $G\backslash i \cup x \to y$. This proves claim 1.

We now turn our attention to forming indicator polynomials. Our goal will be to associate a degree $k-1$ multilinear polynomial to $(k+1)$-sets of rows. We will show the polynomials, are independent, hence the number of rows chosen will be $O(m^{k-1})$.

Let $G = \{i_1, i_2, \ldots, i_{k-1}\}$ and let $G \to i_k, G \to i_{k_1} \in \mathcal{O}'$. Set $S = \{i_1, i_2, \ldots, i_{k+1}\}$. To simplify the notation, we take $i_j = j$ so that $S = [k+1]$. For each $i \in S$, define $S_i = S \setminus i$. By Proposition **??**, on each set S_i, there is some column in short supply. For $i = k, k+1$, this column is necessarily $(0, 0, \ldots, 0, 1)^T$. For a general $i \in S$, we will write a column in short supply as

$$y_i = (a_{1,i}, a_{2,i}, \ldots, a_{i-1,i}, a_{i+1,i}, \ldots, a_{k+1,i})^T.$$

By Proposition **??**, we may chose each y_i for $i \in \{1, 2, \ldots, k-1\}$ so that at least one of $a_{k,i}$ and $a_{k+1,i}$ is equal to 0. We will do so. Now, for $i \in S$, set

$$y_i^0 = (a_{1,i}, a_{2,i}, \ldots, a_{i-1,i}, 0, a_{i+1,i}, \ldots, a_{k+1,i})^T$$

and

$$y_i^1 = (a_{1,i}, a_{2,i}, \ldots, a_{i-1,i}, 1, a_{i+1,i}, \ldots, a_{k+1,i})^T.$$

Notice that $y_i^1 - y_i^0 = e_i$, where e_i is the standard basis vector with a 1 in the i^{th} row. Now choose any $\ell \in S\backslash\{k, k+1\}$, i.e. $\ell \in [k-1]$. We obtain a quite simple dependency:

$$\left(\sum_{j \in S_\ell}^{k} a_{j,\ell}(y_j^1 - y_j^0)\right) - y_\ell^0 = \mathbf{0}. \tag{5}$$

Observe also that $y_k^1 = y_{k+1}^1 = (0, 0, \ldots, 0, 1, 1)^T$. Hence we have the equation

$$y_k^0 + y_{k+1}^0 - y_k^1 = \mathbf{0}. \tag{6}$$

We would like to use these two equations to get a dependence whose sum of coefficients is 0 while having some non-zero entries. Adding these equations together yields the formula

$$\sum_{\substack{j=1\\ j\neq i}}^{k-1} a_{j,\ell} y_j^1 - \sum_{\substack{j=1\\ j\neq i}}^{k-1} a_{j,\ell} y_j^0 - y_i^0 + (1-a_{k,\ell}) y_k^0 + (a_{k,\ell} + a_{k+1,\ell} - 1) y_k^1 + (1 - a_{k+1,\ell}) y_{k+1}^0 = \mathbf{0} \tag{7}$$

We have written the above equation in such a way to show the coefficient in front of each of the columns y_i^ε. We can readily check that the sum of the coefficients is 0. To verify that some coefficients are non-zero recall that y_ℓ was chosen so that at least one of $a_{k,\ell}$and $a_{k+1,\ell}$ are 0 and hence either $(1 - a_{k,\ell}) \neq 0$ or $(1 - a_{k+1,\ell}) \neq 0$ or both. Hence in (**??**), either the coefficient of y_k^0 is non-zero or the coefficient of y_{k+1}^0 is non-zero or both.

For each $i \in S$ and $\varepsilon = 0, 1$ form the new indicator polynomial of (**??**):

$$f_i^\varepsilon = (x_i - \varepsilon) \prod_{j \neq i} (x_j - a_{j,i}).$$

These polynomials are indicator polynomials for the columns y_i^ε in the sense that they identify the (0,1)-complement of the column $(y_i^\varepsilon)^c$.

$$f_i^\varepsilon(\gamma^c) = \begin{cases} (-1)^a & \text{where } a = \sigma_1(\gamma) \text{ and } \gamma|_S = y_i^\varepsilon \\ 0 & \text{otherwise} \end{cases}$$

We use the indicator polynomials for the complements of columns in order to apply Lemma **??**. Now form the polynomial

$$p = \sum_{\substack{j=1\\ j\neq \ell}}^{k-1} a_{j,\ell} f_j^1 - \sum_{\substack{j=1\\ j\neq \ell}}^{k-1} a_{j,\ell} f_j^0 - f_i^0 + (1 - a_{k,\ell}) f_k^0 + (a_{k,\ell} + a_{k+1,\ell} - 1) f_k^1 + (1 - a_{k+1,\ell}) f_{k+1}^0.$$

Note the similarity between the definition of p and the equation (**??**).

Claim 2. The polynomial p has degree at most $k - 1$.

This now follows immediately from Lemma **??**.

Claim 3. For $\{u, v\} = \{k, k+1\}$ appropriately paired, the polynomial p is non-zero evaluated on columns violating $G \to u$ but not violating $G \to v$.

By our choice of ℓ, at least one of $a_{k,\ell}$ or $a_{k+1,\ell}$ is equal to 0. Without loss of generality, suppose that $a_{k+1,\ell} = 0$, and in this case we let $u = k$ and $v = k+1$. Consider the column $y_k^0 = (0, 0, \dots, 1, 0)^T$, which violates $G \to u$ but not $G \to v$. Notice that

$$f_i^\varepsilon((y_k^0)^c) = \begin{cases} (-1)^{k+1} & \text{if } i = k \text{ and } \varepsilon = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the only non-zero term in $p((y_k^0)^c)$ is $(1 - a_{k,\ell})(-1)^{k+1}$, and so the polynomial p is non-zero on the column $(y_k^0)^c$. This gives the claim.

In this manner, we have found a degree $k-1$ polynomial $p = p_S$ which is 0 on γ^c where $\gamma \neq y_i^\varepsilon$, and possibly non-zero when $\gamma = y_i^\varepsilon$. As each y_i is in short supply, there are at most $(k+1)t$ columns where $p(\gamma^c) \neq 0$.

Claim 4. We can delete $O(m^{k-1})$ columns from A to obtain A' so that for every $G \in \binom{[m]}{k-1}$ for which there are any outer implications $G \to i \in \mathcal{O}'$ one can choose an outer implication $G \to t \in \mathcal{O}'$ so that if a column of A' violates any outer implication $G \to i \in \mathcal{O}'$ then it also violates the implication $G \to t$.

Let $\mathcal{F} = \{G \cup \{i_k, i_{k+1}\} \in \binom{[m]}{k+1} \mid G \to i_k, G \to i_{k+1} \in \mathcal{O}'\}$. Greedily choose sets $S_1, S_2, \dots, S_r \in \mathcal{F}$ and columns $\gamma_1, \gamma_2, \dots, \gamma_r$ so that $p_{S_i}(\gamma_i^c) \neq 0$, and $p_{S_j}(\gamma_i^c) = 0$ for all $j < i$. Create an evaluation matrix $E = (e_{ij})$ with entries $e_{ij} = p_{S_i}({\gamma_j}^c)$. Observe that this matrix is upper triangular with non-zero diagonal entries, and hence the chosen polynomials p_{S_i} are linearly independent. As all the polynomials p_S are multilinear of degree at most $k-1$, we have chosen at most $\binom{m}{k-1} + \binom{m}{k-2} + \cdots + \binom{m}{0}$ polynomials, sets and columns. Furthermore, at each S_i at most $(k+1)t$ columns of A have violations. We delete there $O(m^{k-1})$ columns from A to obtain A'.

After deleting columns, consider a fixed $G \in \binom{[m]}{k-1}$. Form a digraph on rows i for which $G \to i \in \mathcal{O}'$, and add the arc (i,j) if every column of A' with a violation of $G \to i$ also violates $G \to j$. We see that this digraph is transitive. Remarkably this digraph is also complete. If the digraph is empty or has one vertex, this is vacuously true. If the digraph has two vertices i and j, let $S = G \cup \{i,j\}$. We apply Claim 2, and with $u = i$ and $v = j$ in some order, the polynomial p_S is non-zero on columns violating $G \to u$ but not violating $G \to v$. In other words, p_S is non-zero on any column

$$\begin{array}{r} G\left\{\begin{array}{l} \\ \\ \\ \end{array}\right. \\ u \\ v \end{array} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

But we have already deleted all columns with $p_S(\overline{\gamma}) \neq 0$. Hence, if any remaining column violates $G \to u$, it must be of the form

$$\begin{array}{c} G\left\{\begin{array}{c} \\ \\ \\ \end{array}\right. \\ u \\ v \end{array} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

We conclude that any column that violates $G \to u$ also violates $G \to v$, so there will be an edge, in some direction, joining the vertices u and v. This verifies that the digraph is a complete transitive digraph. We now choose the terminal vertex t in our graph, and by simple induction, any column violating $G \to i$ will also violate $G \to t$. This proves Claim 4.

We are now basically done. For each $G \in \binom{[m]}{k-1}$ we remove any columns violating $G \to t$, as in the preceeding claim. There are at most $t\binom{m}{k-1}$ columns, namely $O(m^{k-1})$. No more outer violations remain. Call the resulting matrix A''.

Certainly, the matrix A'' contains no violations of outer implications $G \to i \in \mathcal{O}'$, but recall from the beginning of the proof of this lemma that $\mathcal{O}'$ is a subset of all outer implications $\mathcal{O}$. Further recall that we deleted an outer implication from $\mathcal{O}$ if, whenever it was violated, some remaining violation was also violated. Hence, if some outer violation $G \to i \in \mathcal{O} \setminus \mathcal{O}'$ is still violated at column γ in A'', then the column γ must violate some implication in $\mathcal{O}'$. But this is what we have just proved does not exist in A'', hence there are no outer violations in A''.

We have shown that it suffices to delete $O(m^{k-1})$ columns to remove all violations of outer implications. $\square$

After all this work, we are finally ready to prove Theorem **??**.

Proof of Theorem **??**. Let A be an $m \times n$ matrix avoiding F_k. Let $\mathcal{T}$ be the set of rows on which there are two or more columns in short supply. Apply Theorem **??** to remove $O(m^{k-1})$ columns from A, forming a new matrix A' on which there is a columns missing from every k-set $T \in \mathcal{T}$.

Because we are dealing with F_k, there may still be some k-sets of rows where no column is missing. Call this collection of rows $\mathcal{S}$. By Proposition **??**, this column is necessarily $(0, 0, \ldots, 0, 1)^T$, and hence all violations in $\mathcal{S}$ correspond to either inner or outer implications. Using Lemma **??** on A', we can remove $O(m^{k-1})$ columns of A' to form a matrix A'' with no inner implications. Applying Lemma **??** to A'', we delete $O(m^{k-1})$ columns of A'' to form a new matrix A''' with no outer violations. After all these changes, the column $(0, 0, \ldots, 0, 1)^T$ in some ordering will be absent on every k-set $S \in \mathcal{S}$. By the deletions done on A to produce A' we know that for every $S \in 2^{[m]} \backslash \mathcal{S}$ that there is some column missing. We have thus created a matrix A''' for which some column is absent on every k-set of rows. Hence A''' avoids K_k, and by Theorem **??** A''' contains $O(m^{k-1})$ columns. As we have only removed a total of $O(m^{k-1})$ columns, our original matrix A has only $O(m^{k-1})$ columns, as desired. $\square$

We hope that this essay will enable the gentle reader to find applications of linear algebra in their research.

References

[1] N. Alon, On the Density of Sets of Vectors. *Discrete Math* **46** (1983), 199-202.

[2] R.P. Anstee, General forbidden configuration theorems, Journal of Combinatorial Theory Ser. A 40 (1985), 108-124.

[3] R.P. Anstee, A Survey of Forbidden Configurations results, http://www.math.ubc.ca/∼ anstee.

[4] R.P. Anstee, B. Fleming, Linear Algebra Methods for Forbidden Configurations, Combinatorica, to appear

[5] R.P. Anstee, B. Fleming, Two refinements of the bound of Sauer, Perles and Shelah and Vapnik and Chervonenkis. *Discrete Mathematics* **310** (2010) 3318-3323.

[6] R.P. Anstee, B. Fleming, Z. Füredi, and A. Sali, colour critical hypergraphs and forbidden configurations, proceedings of EuroComb 2005, Berlin, Germany. *Discrete mathematics and Theoretical Computer Science*, 2005, 117-122.

[7] R.P. Anstee, General Forbidden Configuration Theorems, *J. Combin. Th. Ser A* **40**(1985), 108-124.

[8] R.P. Anstee, A Forbidden Configuration Theorem of Alon, *J. Combin. Th. Ser A* **47**(1988) 16-27.

[9] R.P. Anstee, A. Sali, Small forbidden configurations IV. *Combinatorica* **25** (2005), 503-518.

[10] A.J. Bondy, U.S.R. Murty,. (2008). *Graph Theory.* London: Springer-Verlag.

[11] C.J. Colbourn, J.H. Dinitz, CRC Handbook of Combinatorial Designs. *CRC Press* (1996), 47-66.

[12] P. Erdős, C. Ko, R. Rado, Intersection theorems for finite sets, *Quarterly Journal of Mathematics.* **12**(1961), 313-320.

[13] P. Frankl, J. Pach, On the number of sets in an null t-design, *European J. Combin.* **4**(1983), 21-23.

[14] P. Keevash, The existence of Designs, arXiv:1401.3665, 2014.

[15] H.J. Ryser, A fundamental matrix equation for finite sets. *Proc. Math. Soc.* **34** (1972) 332-336.

[16] H.J. Ryser, Intersection Properties of Finite Sets. *J. Combin. Th. Ser A* **14**(1973), 79-92.

[17] N. Sauer, On the density of families of sets, *J. Combin. Th. Ser A* **13**(1972), 145-147.

[18] S. Shelah, A combinatorial problem: Stability and order for models and theories in infinitary languages, *Pac. J. Math.* **4**(1972), 247-261.

[19] R. Smolensky, Well known bound for VC-dimension made easy, *Computational Complexity* **6** (1997) 299-300.

[20] E. Sperner, Ein Satz über Untermengen einer endlichen Menge. *Mathematische Zeitschrift* **27** (1928) 544-548.

[21] V.N. Vapnik, A.Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Th. Prob. and Applics.* **16**(1971), 264-280.